Combined Adaptive Filter with LMS-Based Algorithms

Božo Krstajić, LJubiša Stanković, Zdravko Uskoković

Abstract— **A combined adaptive filter is proposed. It consists of parallel LMS-based adaptive FIR filters and an algorithm for choosing the better among them. As a criterion for comparison of the considered algorithms in the proposed filter, we take the ratio between bias and variance of the weighting coefficients. Simulations results confirm the advantages of the proposed adaptive filter.**

I. Introduction

Adaptive filters have been applied in signal processing and control, as well as in many practical problems, [1], [2]. Performance of an adaptive filter depends mainly on the algorithm used for updating the filter weighting coefficients. The most commonly used adaptive systems are those based on the Least Mean Square (LMS) adaptive algorithm and its modifications (LMS-based algorithms).

The LMS is simple for implementation and robust in a number of applications [1], [2], [3]. However, since it does not always converge in an acceptable manner, there have been many attempts to improve its performance by the appropriate modifications: sign algorithm (SA) [8], geometric mean LMS (GLMS) [5], variable step-size LMS (VS LMS) [6], [7].

Each of the LMS-based algorithms has at least one parameter that should be defined prior to the adaptation procedure (step for LMS and SA; step and smoothing coefficients for GLMS; various parameters affecting the step for VS LMS). These parameters crucially influence the filter output during two adaptation phases: transient and steady state. Choice of these parameters is mostly based on some kind of trade-off between the quality of algorithm performance in the mentioned adaptation phases.

AEUE International Journal of Electronics and Communications, Vol.57, No.4, 2003.

We propose a possible approach for the LMS-based adaptive filter performance improvement. Namely, we make a combination of several LMS-based FIR filters with different parameters, and provide the criterion for choosing the most suitable algorithm for different adaptation phases. This method may be applied to all the LMS-based algorithms, although we here consider only several of them.

The paper is organized as follows. An overview of the considered LMS-based algorithms is given in Section 2. Section 3 proposes the criterion for evaluation and combination of adaptive algorithms. Simulation results are presented in Section 4.

II. LMS based algorithms

Let us define the input signal vector $\bar{X}_k = [x(k)\, x(k-1) ... x(k-N+1)]^T$ and vector of weighting coefficients as $\bar{W}_k = [W_0(k)\, W_1(k) ... W_{N-1}(k)]^T$. The weighting coefficients vector should be calculated according to:

$$\overline{W}_{k+1} = \overline{W}_k + 2\mu E\{e_k \overline{X}_k\} \tag{1}$$

where μ is the algorithm step, $E\{\cdot\}$ is the estimate of the expected value and $e_k = d_k - \overline{W}_k^T \overline{X}_k$ is the error at the instant k, and d_k is a reference signal. Depending on the estimation of expected value in (1), one defines various forms of adaptive algorithms: the LMS ($E\{e_k \overline{X}_k\} = e_k \overline{X}_k$), the GLMS ($E\{e_k \overline{X}_k\} = a \sum_{i=0}^{k} (1-a)^i e_{k-i} \overline{X}_{k-i}, 0 < a \le 1$), and the SA ($E\{e_k \overline{X}_k\} = \overline{X}_k\, sign(e_k)$), [1], [2], [5], [8]. The VS LMS has the same form as the LMS, but in the adaptation the step $\mu(k)$ is changed [6], [7].

The considered adaptive filtering problem consists in trying to adjust a set of weighting coefficients so that the system output, $y_k = \overline{W}_k^T \overline{X}_k$, tracks a reference signal, assumed as

$d_k = \overline{W}_k^{*T}\overline{X}_k + n_k$, where n_k is a zero mean Gaussian noise with the variance σ_n^2, and $\overline{W}_k^*$ is the optimal weight vector (Wiener vector). Two cases will be considered: $\overline{W}_k^* = \overline{W}^*$ is a constant (stationary case) and $\overline{W}_k^*$ is time-varying (nonstationary case). In nonstationary case the unknown system parameters (i.e. the optimal vector $\overline{W}_k^*$) are time variant. It is often assumed that variation of $\overline{W}_k^*$ may be modeled as $\overline{W}_{k+1}^* = \overline{W}_k^* + \bar{Z}_k$,where $\bar{Z}_k$ is the zero-mean random perturbation, independent on $\overline{X_k}$ and n_k, with the autocorrelation matrix $G = E\left[\bar{Z}_k\overline{Z}_k^T\right] = \sigma_Z^2\overline{I}$. Note that analysis for the stationary case directly follows for $\sigma_Z^2 = 0$. The weighting coefficient vector converges to the Wiener one, if the condition from [1], [2] is satisfied.

Define the weighting coefficients misalignment, [1], [2], [3], $\overline{V}_k = \overline{W}_k - \overline{W}_k^*$. It is due to both the effects of gradient noise (weighting coefficients variations around the average value) and the weighting vector lag (difference between the average and the optimal value), [3]. It can be expressed as:

$$\overline{V}_k = (\overline{W}_k - E(\overline{W}_k)) + (E(\overline{W}_k) - \overline{W}_k^*), \quad (2)$$

According to (2), the i-th element of $\overline{V}_k$ is:

$$\begin{aligned} V_i(k) &= (E(W_i(k)) - W_i^*(k)) \\ &\quad + (W_i(k) - E(W_i(k))) \\ &= bias(W_i(k)) + \rho_i(k) \qquad (3) \end{aligned}$$

where $bias(W_i(k))$ is the weighting coefficient bias and $\rho_i(k)$ is a zero-mean random variable with the variance σ^2. The variance depends on the type of LMS-based algorithm, as well as on the external noise variance σ_n^2. Thus, if the noise variance is constant or slowly-varying, σ^2 is time invariant for a particular LMS-based algorithm. In that sense, in the analysis that follows we will assume that σ^2 depends only on the algorithm type, i.e. on its parameters.

An important performance measure for an adaptive filter is its mean square deviation (MSD) of weighting coefficients. For the adaptive filters, it is given by, [3]: $MSD = \lim_{k\to\infty} E\left[\bar{V}_k^T\overline{V}_k\right]$.

III. Combined adaptive filter

The basic idea of the combined adaptive filter lies in parallel implementation of two or more adaptive LMS-based algorithms, with the choice of the best among them in each iteration [9]. Choice of the most appropriate algorithm, in each iteration, reduces to the choice of the best value for the weighting coefficients. The best weighting coefficient is the one that is, at a given instant, the closest to the corresponding value of the Wiener vector.

Let $W_i(k,q)$ be the $i-th$ weighting coefficient for LMS-based algorithm with the chosen parameter q at an instant k. Note that one may now treat all the algorithms in a unified way (LMS: $q \equiv \mu$, GLMS: $q \equiv a$, SA: $q \equiv \mu$). LMS-based algorithm behavior is crucially dependent on q. In each iteration there is an optimal value q_{opt}, producing the best performance of the adaptive algorithm. Analyze now a combined adaptive filter, with several LMS-based algorithms of the same type, but with different parameter q.

The weighting coefficients are random variables distributed around the $W_i^*(k)$, with $bias(W_i(k,q))$ and the variance σ_q^2, related by [4], [9]:

$$|W_i(k,q) - W_i^*(k) - bias(W_i(k,q))| \leq \kappa\sigma_q, \quad (4)$$

where (4) holds with the probability $P(\kappa)$, dependent on κ. For example, for $\kappa = 2$ and a Gaussian distribution, $P(\kappa) = 0.95$ (two sigma rule).

Define the confidence intervals for $W_i(k,q)$, [4], [9]:

$$D_i(k) = [W_i(k,q) - 2\kappa\sigma_q, W_i(k,q) + 2\kappa\sigma_q] \quad (5)$$

Then, from (4) and (5) we conclude that, as long as $|bias(W_i(k,q))| < \kappa\sigma_q$, $W_i^*(k) \in D_i(k)$, independently on q. This means that, for small bias, the confidence intervals, for different $q's$ of the same LMS-based algorithm, intersect. When, on the other hand, the bias becomes large, then the central positions of the intervals for different $q's$ are far apart, and they do not intersect.

Since we do not have apriori information about the $bias(W_i(k,q))$, we will use a specific statistical approach to get the criterion for the choice of adaptive algorithm, i.e. for the values of q. The criterion follows from the trade-off condition that bias and variance are of the same order of magnitude, i.e. $|bias(W_i(k,q))| \cong \kappa\sigma_q$, [4].

The proposed combined algorithm (CA) can now be summarized in the following steps:

Step 1. Calculate $W_i(k,q)$ for the algorithms with different $q's$ from the predefined set $Q = \{q_1, q_2, ...\}$.

Step 2. Estimate the variance σ_q^2 for each considered algorithm.

Step 3. Check if $D_i(k)$ intersect for the considered algorithms. Start from an algorithm with largest value of variance, and go toward the ones with smaller values of variances. According to (4), (5) and the trade-off criterion, this check reduces to the check if

$$|W_i(k,q_m) - W_i(k,q_l)| < 2\kappa(\sigma_{q_m} + \sigma_{q_l}) \quad (6)$$

is satisfied, where $q_m, q_l \in Q$, and the following relation holds: $\forall q_h$: $\sigma_{q_m}^2 > \sigma_{q_h}^2 > \sigma_{q_l}^2, \Rightarrow q_h \notin Q$.

If no $D_i(k)$ intersect (large bias) choose the algorithm with largest value of variance. If the $D_i(k)$ intersect, the bias is already small. So, check a new pair of weighting coefficients or, if that is the last pair, just choose the algorithm with the smallest variance. First two intervals that do not intersect mean that the proposed trade-off criterion is achieved, and choose the algorithm with large variance.

Step 4. Go to the next instant of time.

The smallest number of elements of the set Q is $L = 2$. In that case, one of the $q's$ should provide good tracking of rapid variations (the largest variance), while the other should provide small variance in the steady state. Observe that by adding few more $q's$ between these two extremes, one may slightly improve the transient behavior of the algorithm.

Note that the only unknown values in (6) are the variances. In our simulations we estimate σ_q^2 as in [4]:

$$\sigma_q = median(|W_i(k) - W_i(k-1)|)/0.675\sqrt{2}, \quad (7)$$

for $k = 1, 2, .., L$ and $\sigma_Z^2 \ll \sigma_q^2$.

The alternative way is to estimate σ_n^2 as:

$$\sigma_n^2 \approx \frac{1}{T}\sum_{i=1}^{T} e_i^2, \;\; for \;\; x(i) = 0. \quad (8)$$

Expressions relating σ_n^2 and σ_q^2 in steady state, for different types of LMS-based algorithms, are known from literature. For the standard LMS algorithm in steady state, σ_n^2 and σ_q^2 are related by $\sigma_q^2 = q\sigma_n^2$, [3]. Note that any other estimation of σ_q^2 is valid for the proposed filter.

Complexity of the CA depends on the constituent algorithms (Step 1), and on the decision algorithm (Step 3). Calculation of weighting coefficients for parallel algorithms does not increase the calculation time, since it is performed by a parallel hardware realization, thus increasing the hardware requirements. The variance estimations (Step 2), negligibly contribute to the increase of algorithm complexity, because they are performed at the very beginning of adaptation and they are using separate hardware realizations. Simple analysis shows that the CA increases the number of operations for, at most, $N(L-1)$ additions and $N(L-1)$ IF decisions, and needs some additional hardware with respect to the constituent algorithms.

A. Illustration of combined adaptive filter

Consider a system identification by the combination of two LMS algorithms with different steps. Here, the parameter q is μ, i.e. $Q = \{q_1, q_2\} = \{\mu, \mu/10\}$.

The unknown system has four time-invariant coefficients, and the FIR filters are with $N = 4$. We give the average mean square deviation ($AMSD$) for both individual algorithms, as well as for their combination, Fig. 1. Results are obtained by averaging over 100 independent runs (the Monte Carlo method), with $\mu = 0.1$. The reference d_k is corrupted by a zero-mean uncorrelated Gaussian noise with $\sigma_n^2 = 0.01$ and $SNR = 15\ dB$, and κ is 1.75. In the first 30 iterations the variance was estimated according to (7), and the CA picked the weighting coefficients calculated by the LMS with μ.

As presented in Fig. 1, the CA first uses the LMS with μ and then, in the steady state, the LMS with $\mu/10$. Note the region, between the 200*th* and 400*th* iteration, where the algorithm can take the LMS with either step-size, in different realizations. Here, performance of the CA would be improved by increasing the number of parallel LMS algorithms with steps between these two extrems. Observe also that, in steady state, the CA does not ideally pick up the LMS with smaller step. The reason is in the statistical nature of the approach.

Combined adaptive filter achieves even better performance if the individual algorithms, instead of starting an iteration with the coefficient values taken from their previous iteration, take the ones chosen by the CA. Namely, if the CA chooses, in the k-th iteration, the weighting coefficient vector $\bar{W}_p$, then each individual algorithm calculates its weighting coefficients in the $(k+1)$-th iteration according to:

$$\overline{W}_{k+1} = \overline{W_p} + 2\mu E\{e_k \overline{X_k}\} \qquad (9)$$

Fig. 2 shows this improvement, applied on the previous example. In order to clearly compare the obtained results, for each simulation we calculated the $AMSD$. For the first LMS (μ) it was $AMSD$= 0.02865, for the second LMS ($\mu/10$) it was $AMSD$= 0.20723, for the CA (CoLMS) it was $AMSD$= 0.02720 and for the CA with modification (9) it was $AMSD$=0.02371.

IV. Simulation Results

The proposed combined adaptive filter with various types of LMS-based algorithms is implemented for stationary and nonstationary cases in a system identification setup. Performance of the combined filter is compared with the individual ones, that compose the particular combination.

In all simulations presented here, the reference d_k is corrupted by a zero-mean uncorrelated Gaussian noise with $\sigma_n^2 = 0.1$ and $SNR = 15\ dB$. Results are obtained by averaging over 100 independent runs, with $N = 4$, as in the previous section.

(a) **Time Varying Optimal Weighting Vector**: The proposed idea may be applied to the SA algorithms in a nonstationary case. In the simulation, the combined filter is composed out of three SA adaptive filters with different steps, i.e. $Q = \{\mu,\ \mu/2,\ \mu/8\}$; $\mu = 0.2$. The optimal vector is generated according to the presented model with $\sigma_Z^2 = 0.001$, and with $\kappa = 2$. In the first 30 iterations the variance was estimated according to (7), and CA takes the coefficients of SA with μ (SA1).

Figure 3 shows the $AMSD$ characteristics for each algorithm. In steady state the CA does not ideally follow the SA3 with $\mu/8$, because of the nonstationary problem nature and a relatively small difference between the coefficient variances of the SA2 and SA3. However, this does not affect the overall performance of the proposed algorithm. $AMSD$ for each considered algorithm was: $AMSD$= 0.4129 (SA1,μ), $AMSD$=0.4257 (SA2, $\mu/2$), $AMSD$=1.6011 (SA3, $\mu/8$) and $AMSD$=0.2696 (Comb).

(b) Comparison with VS LMS algorithm [6]: In this simulation we take the improved CA (9) from 3.1, and compare its performance with the VS LMS algorithm [6], in the case of abrupt changes of optimal vector. Since the considered VS LMS algorithm [6] updates its step size for each weighting coefficient individually, the comparison of these two algorithms is meaningful.

All the parameters for the improved CA are the same as in 3.1. For the VS LMS algorithm [6], the relevant parameter values are the counter of sign change $m_0 = 11$, and the counter of sign continuity $m_1 = 7$. Figure 4 shows the $AMSD$ for the compared algorithms, where one can observe the favorable properties of the CA, especially after the abrupt changes. Note that abrupt changes are generated by multiplying all the system coefficients by -1 at the 2000-th iteration (Fig. 4). The $AMSD$ for the VS LMS was $AMSD = 0.0425$, while its value for the CA (Co LMS) was $AMSD = 0.0323$.

For a complete comparison of these algorithms we consider now their calculation complexity, expressed by the respective increase in number of operations with respect to the LMS algorithm. The CA increases the number of requres operations for N additions and N IF decisions. For the VS LMS algorithm, the respective increase is: $3N$ multiplications, N additions, and at least $2N$ IF decisions. These values show the advantage of the CA

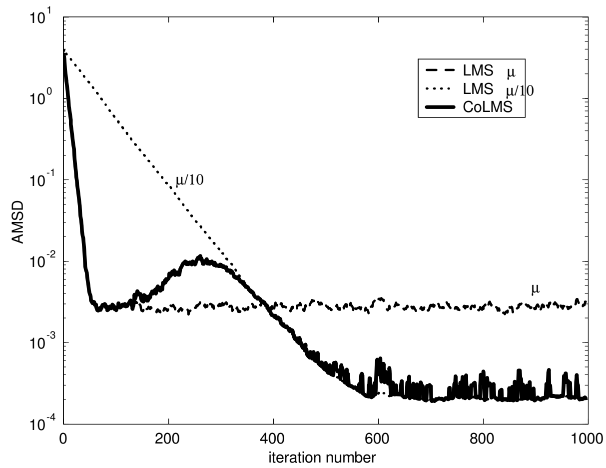

Fig. 1. Average MSD for considered algorithms

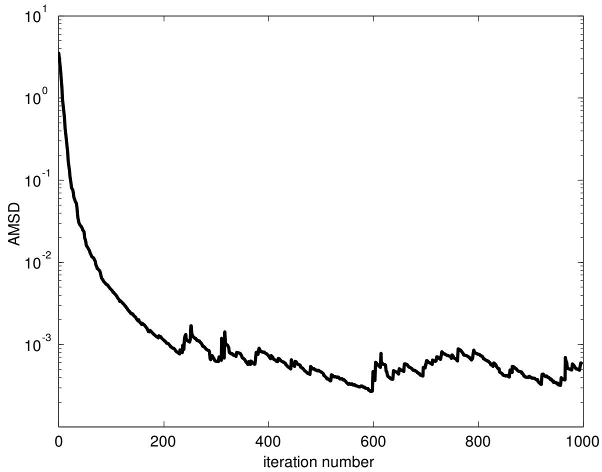

Fig. 2. Average MSD for considered algorithms

with respect to the calculation complexity.

V. Conclusion

Combination of the LMS based algorithms, which results in an adaptive system that takes the favorable properties of these algorithms in tracking parameter variations, is proposed. In the course of adaptation procedure it chooses better algorithms, all the way to the steady state when it takes the algorithm with the smallest variance of the weighting coefficient deviations from the optimal value.

VI. Acknowledgment

This work is supported by the Volkswagen Stiftung, Federal Republic of Germany.

References

[1] B. Widrow and S. Stearns, "Adaptive Signal Processing" , Prentice-Hall, Inc. Englewood Cliffs, N.J. 07632.
[2] S.T. Alexander, "Adaptive Signal Processing - Theory and Applications", Springer-Verlag, New York, 1986.

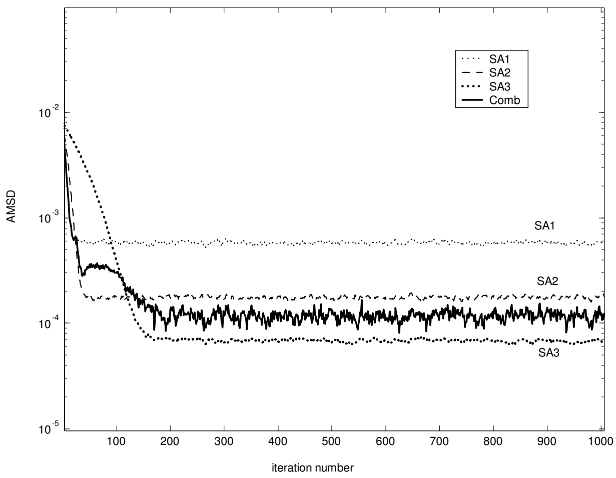

Fig. 3. Average MSD for considered algorithms

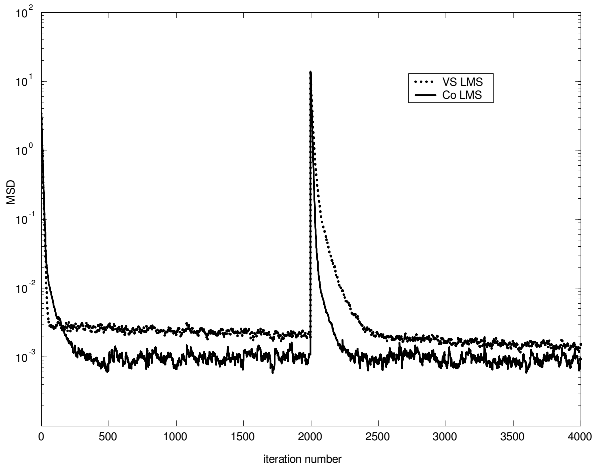

Fig. 4. Average MSD for considered algorithms

[3] B. Widrow, J. M. McCool, M. G. Larimore and C. R. Johnson, "Stationary and Nonstationary Learning Characteristics of the LMS Adaptive Filter", *Proc. IEEE*, Vol. 64, pp. 1151-1161, 1976.

[4] LJ. Stankovic, V. Katkovnik: "Algorithm for the instantaneous frequency estimation using time-frequency distributions with variable window width", *IEEE SP. Letters*, vol. 5, No. 9, Sept. 1998, pp. 224-227.

[5] B. Krstajic, Z. Uskokovic, Lj. Stankovic, "GLMS Adaptive Algorithm in Linear Prediction", *Proc. CCECE'97,* vol.I, pp.114-117, Canada, 1997.

[6] R.W. Harris, D.M. Chabries and F. A. Bishop, "A Variable Step (VS) Adaptive Filter Algorithm", *IEEE Trans. ASSP*, vol. 34, no.2, April 1986, pp.309-316

[7] T. Aboulnasr and K. Mayyas, "A Robust Variable Step-Size LMS-Type Algorithm: Analysis and Simulations" *IEEE Trans. SP*, vol. 45, No. 3, pp.631-639, March 1997.

[8] V. J. Mathews and S. H. Cho: "Improved Convergence Analysis of Stochastic Gradient Adaptive Filters Using the Sign Algorithm", *IEEE Trans. ASSP*, vol.35, No. 4, pp. 450-454 April 1987.

[9] B. Krstajic, LJ. Stankovic, Z. Uskokovic, I. Djurovic: "Combined adaptive system for identification of unknown systems with varying parameters in a noisy enviroment", *Proc. IEEE ICECS'99*, Phafos, Cyprus, Sept. 1999.